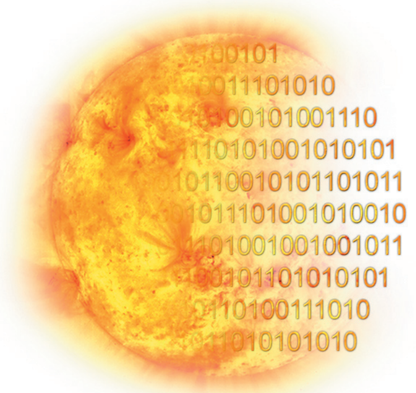
FRBR in a Scientific Data Context



http://www.virtualsolar.org/

J.A. Hourclé (RSIS/NASA-GSFC)
joseph.a.hourcle.1@gsfc.nasa.gov

Abstract

Data can be catalogued at many different levels of granularity – data granules, data products, data sets and data collections. Unfortunately, one discipline's data product is another discipline's data set. The inconsistency of terms creates difficulty in interfacing the archives – if one archive generates metadata records at the data granule level, while another describes their data at the data product level, there will be confusion in merging records from a federated search.

Although the OAIS reference model (CCSDS 2002) discusses the concepts of Collection Descriptions and Representation Networks, it does not discuss the granularity of the data being described, other than as an Archive Information Package (AIP), or an Archive Information Collection (AIC). Unfortunately, the amount of data that makes up an AIC for one archive may be an AIP for another. This poses a problem when archives return different granularity in requests.

Some Active Archives may provide different versions of the same content, be it differing Editions, or alternative packaging for different Designated Communities. Some may return records for Derived AIPs interspersed with their source AIPs. This may be desired by some users, but it can confuse and overwhelm users who are not part of the Local Community for that archive.

Since the scientific community lacks coordination in terminology for these aspects of cataloging, we examine the concepts of the Functional Requirements for Bibliographic Records [FRBR] (München 1998) developed by the library community for best practices in this field. We discuss the applicability of FRBR concepts to scientific data, and the need for a similarly purposed model as the "glue" necessary to hold together any virtual observatory or other federated search system for scientific data.

The Problem

As scientists search for data, they want to see an appropriate level of granularity to select the data that are of interest. Unfortunately, that level can differ depending on the intended use of the data. As federated search tools cross disciplinary boundaries, we cannot assume that a single level of granularity is appropriate for the system's users.

Every level of granularity has its benefits, and so there is no desire to standardize on a specific level, but there is a need to enable archives to easily communicate at what level of detail their data is cataloged.

Equivalence of Records

Without defining the various levels of granularity, it becomes a moot point to attempt to give data persistant identifiers:

Is an identical digital copy of an AIP at another archive equivalent? Not to the scientists, if one is inaccessible to them.

What if the same data and metadata from an AIP are stored in some other format? Not equivalent, if their tools can't handle one of the formats.

What if the SIPs are the same, but the PDI is changed to describe different uses of the data? Not equivalent, if they don't have the proper metadata for the task at hand.

Is the raw data plus the necessary calibration information equivalent to the calibrated data? Not if the scientist is trying to determine if there's something wrong with the calibration.

What is FRBR?

FRBR is an entity-relationship model to provide a structure to discuss data requirements in bibliographic catalogs. FRBR concentrates on four elementary uses of bibliographic records (München 1998).

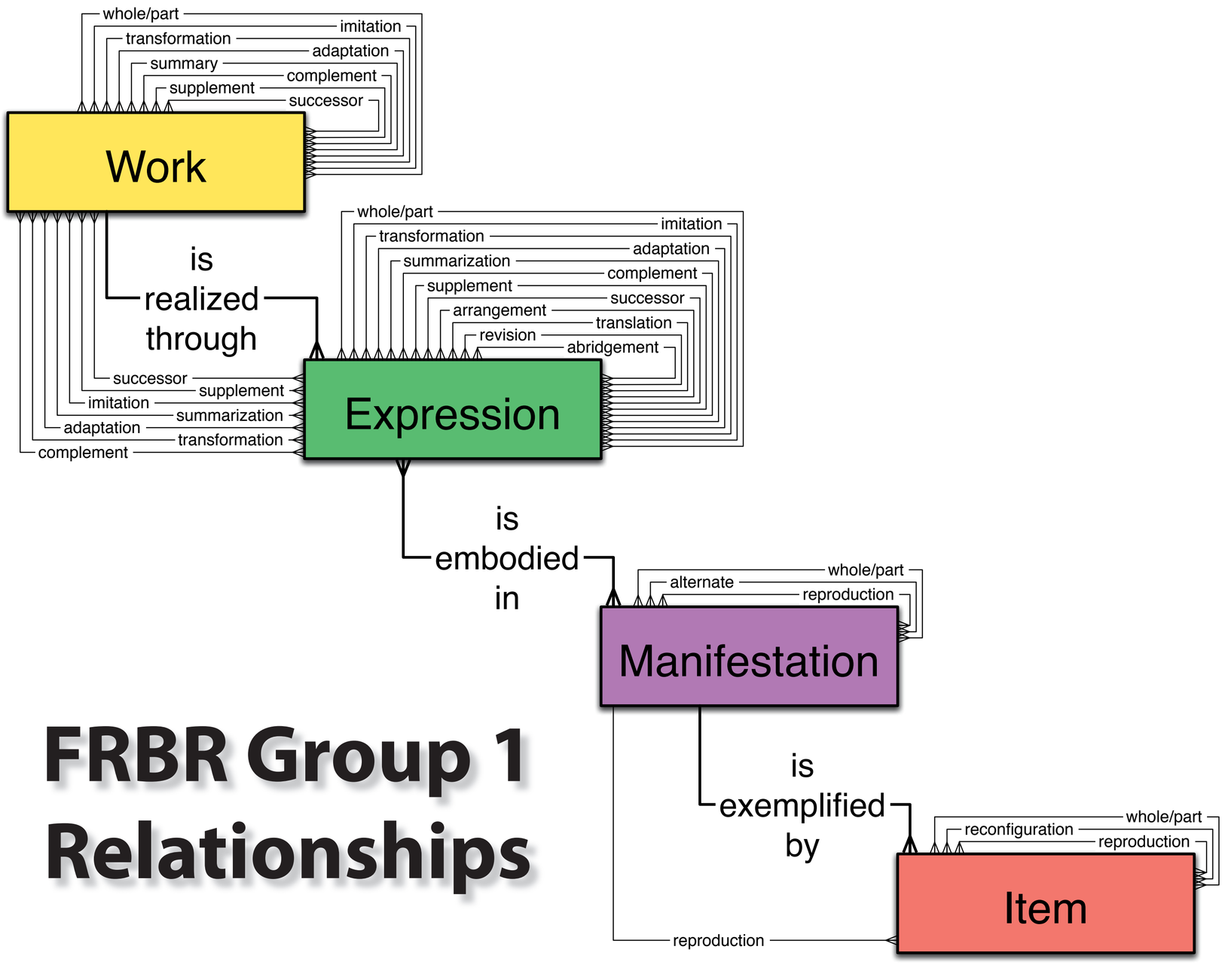
to **find** entities that correspond to the user's stated search criteria

to **identify** an entity (ie, confirm that it corresponds to the entity sought, or to distinguish between multiple similar entities)

to **select** an entity that is appropriate to the user's needs

to acquire or **obtain** access to the entity described

The FRBR tasks correspond to the role of an OAIS Access Aid (Finding, Ordering, Retrieval).



FRBR Group 1 Relationships

How can FRBR help us?

In defining their problem, the library community dealt with the ambiguity of the term 'book'. They broke the polysemous term into four distinct classes:

Work
The abstract story within the book
Expression
The words that tell the story within a book
Manifestation
A specific publication that contains the story
Item
A specific physical instance of a book.

FRBR referrs to these as the 'Group 1' entities. There are also 'Group 2' entities, which are the authors of the works, as well as 'Group 3' entities, which are other possible subjects for the works.

Current work on scientific data models have focused on the subject of data — what is being observed. Although this allows a scientist to search an archive, it does not assist them in selecting the best file to download.

These additional concepts are necessary to provide the user with the specific information they are looking for, in the packaging which they can most readily make use of, from the location that is most useful to them.

Where do we go from here?

FRBR took the library community years to develop. Although we can use it for inspiration, it will require cooperation of space scientists and information scientists to determine what metadata should be standardized in Representation Networks, and how best to share it between archives.

Known Issues with FRBR

FRBR does not provide a specific entity for 'Super Works' or other forms of bibliographic collections (Rosenberg and Hillman, 2004). The LIS community has seen the need to extend the FRBR concepts to better describe serials and other aggregated 'families' of work. The International Federation of Library Associations and Institutes established the Working Group on Aggregates in 2005 (IFLA 2007).

The scientific concept of 'Data Set' most closely corresponds to the bibliographic concept of a serial. Unfortunately, as FRBR does not handle this concept well, there is little chance of using it as a basis for exchanging Collection Descriptions.

References:

CCSDS (2002). "Reference Model for an Open Archival Information System (OAIS)" <http://public.ccsds.org/publications/archive/650x0b1.pdf>.

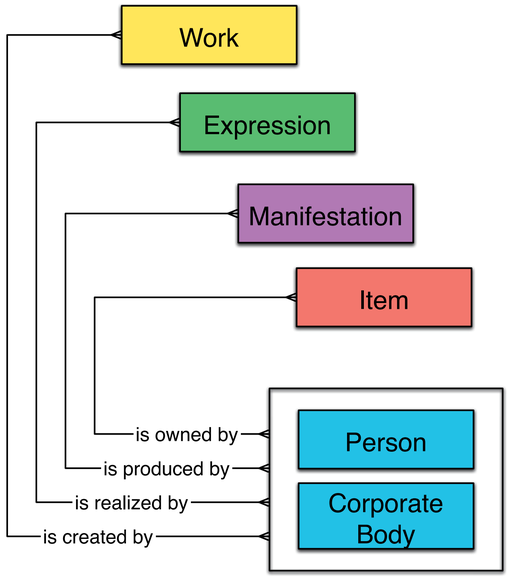
Delsey, T. (2003). "FRBR and Serials" <http://www.ifla.org/VII/s13/wgfrbr/papers/delsey.pdf>.

IFLA (2007). "Working Group on Aggregates" <http://www.ifla.org/VII/s13/wgfrbr/aggregates_wg.htm>.

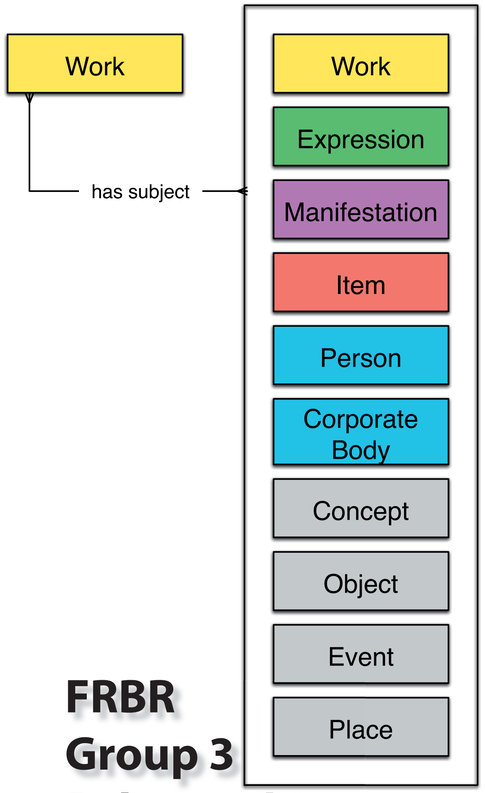
München, K.G.S. (1998). "Functional Requirements for Bibliographic Records, Final Report" <http://www.ifla.org/VII/s13/frbr/frbr.pdf>.

Rosenberg, F. and Hillman, F. (2004). "An Approach to Serials with FRBR in Mind: CONSER Task Force on Universal Holdings," <http://www.lib.unc.edu/cat/mfh/serials_approach_frbr.pdf>.

Tillett, B. (2004). "What is FRBR?: A Conceptual Model for the Bibliographic Universe" <http://www.loc.gov/cds/FRBR.html>.



FRBR Group 2 Relationships



FRBR Group 3 Relationships